The Rationale of Rationalization

Walter Veit, Joe Dewhurst, Krzysztof Dolega, Max Jones, Shaun Stanley, Keith Frankish, and *Daniel C. Dennett*

August 2019

Abstract

Fiery Cushman argues that "[r]ationalization is designed not to accurately infer unconscious mental states, but to construct new ones; it is not a discovery, but a fiction". While we agree in broad strokes with the characterization of rationalization as a 'useful fiction', we think that Cushman's claim remains ambiguous in two crucial respects: (i) the reality of beliefs and desires, i.e. the fictional status of folk psychological entities, and (ii) the degree to which they should be understood as useful and representative. Our aim here is to clarify both points and illuminate how rationalization could be understood as a useful fiction. In doing so, we aim to explicate the Rationale of Rationalization.

KEYWORDS: rationalization, intentional stance, folk psychology, representation

This is an unedited preprint draft for *Behavioral and Brain Sciences*.

Target Article:
Cushman, F. (n.d.). Rationalization is rational. Behavioral and Brain Sciences, 1-69. doi:10.1017/S0140525X19001730

Commentary:
Post-hoc rationalisation, i.e., retrospectively attributing or constructing 'hidden' beliefs and desires inferred from how one has behaved in the past, has traditionally been seen to threaten the idea that humans are 'rational', since it happens subsequent to the process under consideration. If the relevant mental states that are supposed to rationalize an action only come into existence after the action has occurred, then they cannot be treated as the cause of that action. However, Cushman argues that a post-hoc process of this kind can still be seen as 'rational' in the sense that it constructs new beliefs and desires that both serve a useful function and track some underlying adaptive rationales that have shaped the behaviour being rationalized. Rationalization, according to Cushman, is supposed to be a 'useful fiction'. We think that this proposal invites two serious ambiguities, firstly to do with the ontological status of the mental states that are the outputs of rationalisation (i.e., folk psychological states like beliefs and desires), and secondly to do with the degree to which they should be understood as useful and representative. We will address each ambiguity in turn, using our resolution of the latter to help resolve the former.

Throughout his article, Cushman seems to assume a fairly robust understanding of what beliefs and desires are, framing them as functionally discrete internal states with determinate contents. He is committed to the idea that there is a crucial distinction between ‘real’ reasoning processes, which involve operations on beliefs and desires, and the fictional ones produced by rationalization, which don't involve any such operations. Rationalization, on his account, seems to play the role of a process of self-interpretation in which one authors fictions about the causes of one’s own behavior. Drawing these distinctions might not be as easy as Cushman suggests, if there is no principled “dividing line between *genuine* belief-talk or agent-talk and mere *as if* belief-talk and agent-talk” (Dennett 2011, 481). Indeed, the lack of such a dividing line similarly arises for agential descriptions or ‘rationalizations’ in evolutionary biology (see Dennett forthcoming, Veit 2019, Okasha 2018, Tarnita 2017). Without such a dividing line, however, it is unclear what the ontological status of beliefs and desires is supposed to be. If Cushman were to deny that there are anything at all like beliefs and desires prior to the rationalisation process, making the folk psychological states produced by this process entirely fictional, he would fall close to eliminative materialists such as Paul (1981) and Patricia (1986) Churchland. We do not think that Cushman would like to endorse this option, as he seems quite committed to the existence of beliefs and desires. The other option, then, and this is a move we recommend for Cushman, is to commit to the existence of some sort of proto-mental states prior to the rationalisation process, in which case we think it is unclear in what sense the output of the rationalisation process also constitute fictional entities. Of course, the rationalisation process might influence or replace these proto-mental states via a narrative process that we could call ‘fictional’, but it is no longer the mental states themselves that are fictions, rather the process that produces them.

This brings us to the second ambiguity: in what sense can ‘fictional’ mental states (or processes) be understood as useful? Cushman clarifies that these fictions can be useful even when they are not “perfectly accurate representations” by appealing to Dennett’s (1987) ‘intentional stance’, according to which the attribution of beliefs and desires are understood as nothing more than a way of tracking observable patterns in behavior (or the categorical bases of those patterns), and have no further ontological status ‘inside’ the system. However, this comparison reveals a tension in his dual conception of folk-psychological states. Dennett’s intentional stance assumes that habit, instinct, norms etc. may all support rational patterns of behavior, and that this is all that is needed for a system to manifest ‘genuine’ beliefs and desires. It is the fact that these processes support rational responses that makes it worth extracting information from them via rationalization (i.e., by adopting the intentional stance), and then re-presenting this information in a rich belief/desire format. Reformatted in this way, beliefs and desires take the form of the linguistic utterances that Dennett originally called “opinions” (1987) and Frankish has more recently called “superbeliefs” (2004). For us, richness is a matter of having a discrete representational vehicle, such as that provided by natural language, but it is not clear that this is what Cushman has in mind when he talks about beliefs and desires.

As we see it there are two broad ways to achieve such a rich conception of belief, either internal or external. On the internal conception, i.e. traditional computationalism, this vehicle is a neural one, and beliefs are formed and processed at a subpersonal level. On the external conception, the vehicle is natural language, and beliefs are formed and manipulated at a personal level by agents themselves, as a way of describing and regulating their own and others’ behavior. Forming a rich belief, i.e. an opinion or superbelief, is like adopting a policy or making a bet on truth – we commit to taking a sentence as an expression of truth, and regulate our other utterances and commitments accordingly. Cushman seems to espouse a version of the former interpretation, but we think that the latter interpretation is to be preferred, as it can help to resolve the two ambiguities outlined above.

Once this external approach is adopted, the sense in which rationalisation is ‘fictional’ becomes clear: it involves the construction of a narrative that is strictly false with regard to the underlying mechanisms, but nonetheless captures real patterns in the behavior generated by those mechanisms. We propose to interpret rationalisation as the process of taking the austere ‘proto-beliefs' manifested in behaviour, and transforming them into superbeliefs or opinions, (i.e., rich, linguistically formatted beliefs and desires), via the application of the intentional stance to one’s own behavior. Taking this can help to resolve the ambiguities described above, provided that Cushman is willing to adopt this distinction between the austere beliefs that are implicit in all (seemingly) intelligent behavior, and the explicit, linguistically mediated beliefs that are the outcome of the rationalization process. The latter could be seen as ‘fictional’, in the sense that they only came about as the result of a story that we tell about our own behavior, and yet they are also ‘real’, in the sense that they do accurately capture (and help to track) our behavior (even if they do not accurately describe the processes underlying that behavior). By coming to be explicitly represented in natural language, expressing normative commitments, they can also indirectly influence our future behaviour. In short, we think rationalisation should be treated as the reverse engineering of what Dennett (2017) has called “free-floating rationales”, i.e. instinctive behavioral patterns, like avoiding snakes or heights, that are not explicitly encoded but nonetheless ‘make rational sense’. Similarly, the underlying ‘reasons’ that are implicit in our behavior can be inferred (or rather uncovered) via rationalization, which can then lead to further behavioural improvements by engaging in explicit rational deliberation. This is the *Rationale of Rationalization.*

References

Churchland, P.M., 1981, “Eliminative Materialism and the Propositional Attitudes”, *Journal of Philosophy,* 78: 67–90.

Churchland, P.S., 1986, *Neurophilosophy: Toward a Unified Science of the Mind/Brain,* Cambridge, MA: MIT Press.

Cushman, F. (n.d.). „Rationalization is rational“, *Behavioral and Brain Sciences*, 1-69. doi:10.1017/S0140525X19001730

Dennett, D.C. forthcoming. “Clever Evolution”, Review of Samir Okasha’s Agents and Goals in Evolution, Oxford. Oxford University Press, 2018, *Metascience*.

Dennett, D.C. 2011. “Homunculi rule: reflections on Darwinian populations and natural selection by Peter Godfrey-Smith”, *Biology & Philosophy,* 26:475–488.

Dennett, D.C. 1987. *The Intentional Stance,* Cambridge, MA: MIT Press.

Frankish, K. 2004. *Mind and Supermind,* Cambridge University Press.

Okasha, S. 2018. *Agents and Goals in Evolution*. Oxford. Oxford University Press.

Tarnita, C.E. 2017. “The ecology and evolution of social behavior in microbes”, *Journal of Experimental Biology* 2017 220, 18-24.

Veit, W. 2019. “Evolution of multicellularity: cheating done right”, *Biology & Philosophy* 34: 34. Online First. https://doi.org/10.1007/s10539-019-9688-9